

Anvendelse af AI i primær sektoren

Præsentation af model til forudsigelse af kræft baseret på rutinemæssige laboratorietests

Diana Sørensen, Afdelingslæge, Ph.d, Klinisk Biokemisk Afdeling, Sygehus Lillebælt

Jonna Skov Madsen, MD, Ph.d, Professor og Koordinerende Forskningsleder, IRS-SLB/VejleLedende Overlæge, Klinisk Biokemisk Afdeling, Sygehus Lillebælt

Ivan Brandslund, MD, DMScProfessor, Research director, University of Southern Denmark, Klinisk Biokemisk Afdeling, Sygehus Lillebælt

Henry Christensen, Bioinformatiker, Klinisk Biokemisk Afdeling, Sygehus Lillebælt

Søren Gray Worsø Laursen, Kræftens Bekæmpelse

Christian Hardahl, SAS Institute A/S, Danmark



Projekt i et samarbejde

Klinisk Biokemisk Afdeling, Sygehus Lillebælt (SLB)
Syddansk Universitet (SDU)
De praktiserende læger i SLBS optageområde
Røntgen Afdelingen SLB
Kræftens Bekæmpelse
SAS Institut



Formål

- *At vurdere evnen af en AI -model til at give en individuel kræftrisikoscore baseret på rutine analyser anvendt i almen praksis*
- *At sammenligne de risikoscorer, der opnås i AI -modellen, med resultater opnået ved standard logistisk regression (LR)*



Studie population



Siden 2008 har PL haft adgang til at bestille en "profil" med en række laboratorieanalyser som et ekstra værktøj til patienter med uspecifikke symptomer mhp. at identificere personer med øget risiko for kræft



Data grundlaget består af 6592 profiler i perioden november 2011 til Marts 2020



*Data analyser er baseret på:
Kohorte I for udvikling og validering med 5224 profiler
Kohorte II for "out of time" validering med 1368 profiler*



Cohorts

Initial exploration - blood
samples panels

Exclusion criteria

Too few biomarkers

Panels with ≤ 10 laboratory test
results removed

Non adults

Patients < 18 years of age at the time
of the blood sampling excluded

Previous cancer

Patients with a previous cancer
diagnoses within 5 years (5×365 days)
prior to the date for the current blood
sample panels removed

Unknown gender

Cases without information of gender
removed

Repeated patient

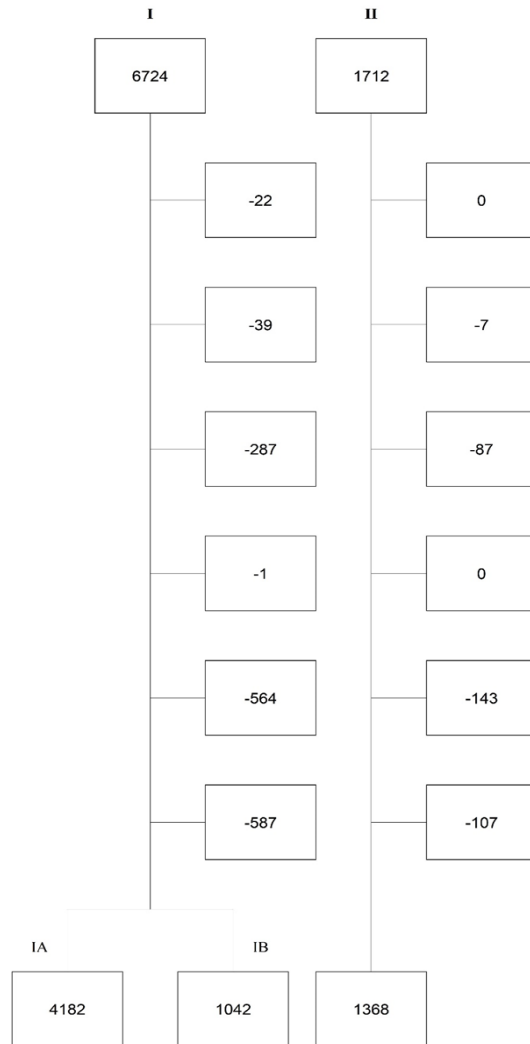
Panels from the patients that had
previously received the blood sample
panel within the last 2 years removed.

Final model

Panels with ≤ 30 laboratory test
results removed

Final cohorts

Cohort IA for training
Cohort IB for validation
Cohort II for test





Blodprøveprofil

- **"Occult/Cancer, M":**

- B-hæmoglobin
- MCV
- MCH
- B-leukocytter med differentialtælling
- B-reticulocytter
- B-trombocytter
- P-C reaktivt protein
- P-natrium
- P-kalium
- P-calcium total
- P-albumin
- P-creatinin
- P-carbamid
- P-urat
- P-glucose
- P-bilirubin
- P-alanintransaminase
- P-basisk phosphatase
- P-amylase pancreasspecifik
- P-laktatdehydrogenase
- P-immunglobuliner (IgA, IgG og IgM)
- P-thyroidea stimulerende hormon
- P-prostataspecifikt antigen

- **"Occult/Cancer, K":**

- B-hæmoglobin
- MCV
- MCH
- B-leukocytter med differentialtælling
- B-reticulocytter
- B-trombocytter
- P-C reaktivt protein
- P-natrium
- P-kalium
- P-calcium total
- P-albumin
- P-creatinin
- P-carbamid
- P-urat
- P-glucose
- P-bilirubin
- P-alanintransaminase
- P-basisk phosphatase
- P-amylase pancreasspecifik
- P-laktatdehydrogenase
- P-immunglobuliner (IgA, IgG og IgM)
- P-thyroidea stimulerende hormon
- P-cancer antigen-125

Metode

- *Datagrundlaget*
- *Outcome*
- *Datastyring og statistik*
- *Bestemmelse af AI –modellerne*



Outcome

- *Registrering: enhver kræft diagnosticeret inden for 730 dage opfølgning efter undersøgelsens inklusion*
- *Alle former for kræft var inkluderet (både primær og metastatisk)*
- *Det vigtigste resultatmål i denne undersøgelse var "kræft inden for 90 dage"*

Metode

- *Datagrundlaget*
- *Outcome*
- *Datastyring og statistik*
- *Bestemmelse af AI –modellerne*



Datastyring og statistik

- *AI -modeludvikling blev udført baseret på kohorte I (opdelt i training og valideringsafsnit(80/20))*
- *Den endelige validering af modellen blev udført i den "out of time", kohorte II.*
- *Alle modelparametre: estimeret i træningsafsnittet; præstationsmål blev hentet fra valideringsafsnittet*
- *Hovedkriterierne for modelvalg var AUC målt på valideringsafsnittet for begge køn*



Metode

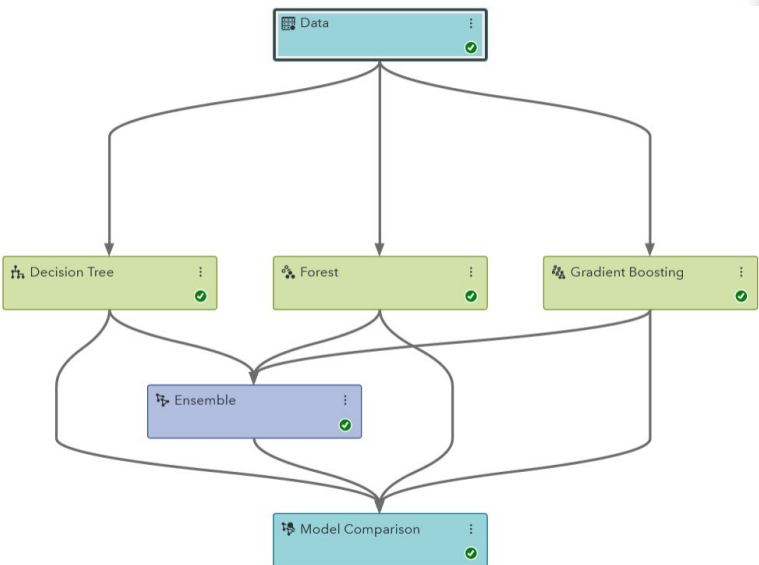
- *Datagrundlaget*
- *Outcome*
- *Datastyring og statistik*
- *Bestemmelse af AI –modellerne*

Bestemmelse af AI –modellerne

- *To indledende strategier for modeludvikling,*
 - *“Random Forrest”metode, der har den fordel, at den ofte tillader gode resultater med et minimum af imputationer og transformationer, da denne model håndterer manglende værdier eksplicit uden at blive påvirket af variabel skala eller størrelse.*
 - *En neural netværksmodel (ANN), som kræver en standardiseringsmetode, men giver mulighed for mere fleksibilitet i modelleringsprocessen.*
 - *Den første modelleringsrunde indikerede en større pasform til ANN - valgt til yderligere finjustering.*

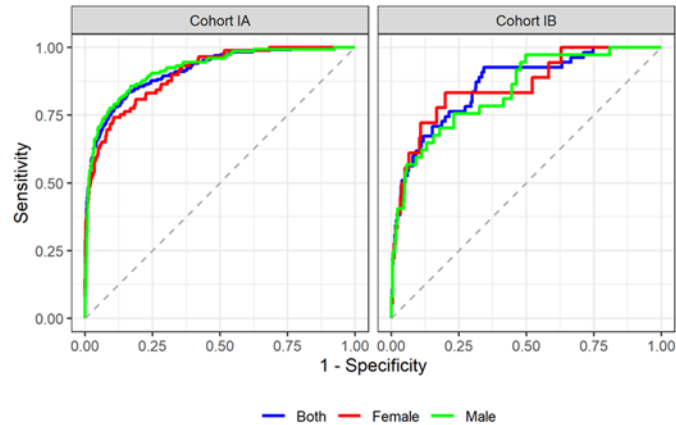


Resultater



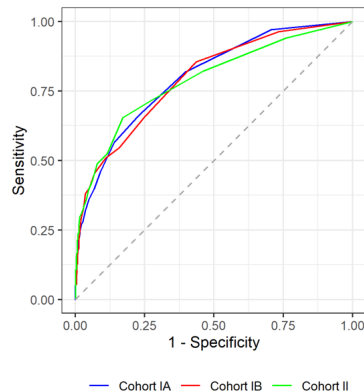
AI model

Kohorte I



AUC ved AI –analysen: 0,91, 0,86 og 0,79 for træningskohort IA, valideringskohort IB og "out of time" valideringstest kohorte II

Logistik regression (LR) analysen



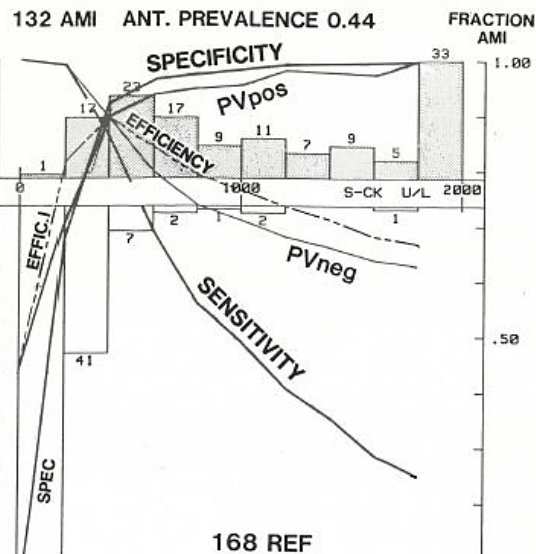
AUC ved LR –analysen: 0,80, 0,80 og 0,79 for træningskohort IA, valideringskohort IB og "out of time" valideringstest kohorte II



Sammenligning af AUC resultater fra AI med LR

- *AUC ved AI –analysen:*
- *0,91 træningskohort IA,*
- *0,86 valideringskohort IB*
- *0,79 "out of time" valideringstest kohorte II*

- *AUC ved LR –analysen:*
- *0,80 træningskohort IA,*
- *0,80 valideringskohort IB*
- *0,79 "out of time" valideringstest kohorte II*



Discr. U/L	Pos Dis	Sens.	Neg Ref	Spec	PVpos	PVneg	Effic. I	Effic. II
0	132	1.00	0	0.00	0.44	-	0.44	1.00
200	131	0.99	114	0.68	0.71	0.99	0.82	1.67
400	114	0.86	155	0.92	0.90	0.90	0.90	1.78
600	91	0.69	162	0.96	0.94	0.80	0.84	1.65
800	74	0.56	164	0.98	0.95	0.74	0.79	1.54
1000	65	0.49	165	0.98	0.96	0.71	0.77	1.47
1200	54	0.41	167	0.99	0.98	0.68	0.74	1.40
1400	47	0.36	167	0.99	0.98	0.66	0.71	1.35
1600	38	0.29	167	0.99	0.97	0.64	0.68	1.28
1800	33	0.25	168	1.00	1.00	0.63	0.67	1.25

Fig 3

Same study as in fig 2. The dependent variables sensitivity, specificity, predictive values, as well as "efficiency I" (see part II of this paper), have been plotted as a function of discriminator position and superimposed together on the histogram. The values of any of these dependent variables may be read directly from the intercepts on the right-hand ordinate. All values are also provided in the table below.

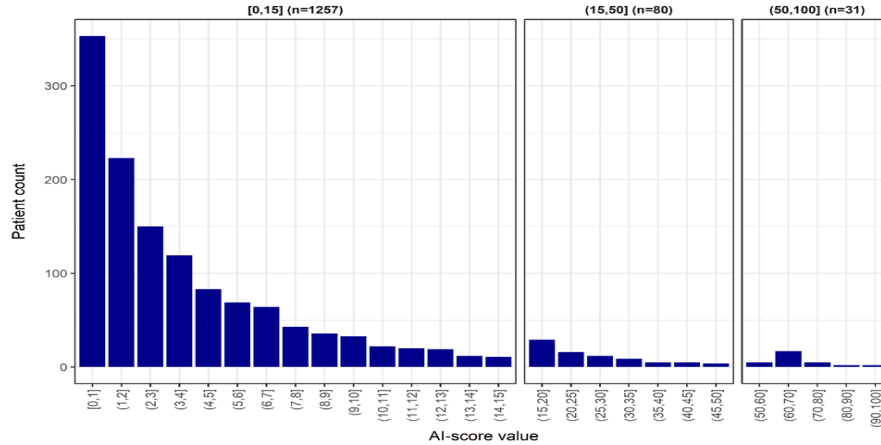
To the evaluator, the consequences of any selected discriminator position are now immediately apparent. Note that maximal "efficiency" (see part two) would indicate a discriminator position at 400 U/L. This, however, would have decreased the sensitivity from 0.99 to 0.86. This, in turn, would have made construction of the diagnostic strategy at suspect AMI impossible (see pp 24-25).

The concepts of Efficiency I and II are discussed in part II of this paper.

AI risikoscore og forekomst af kræft i valideringstest kohorte II

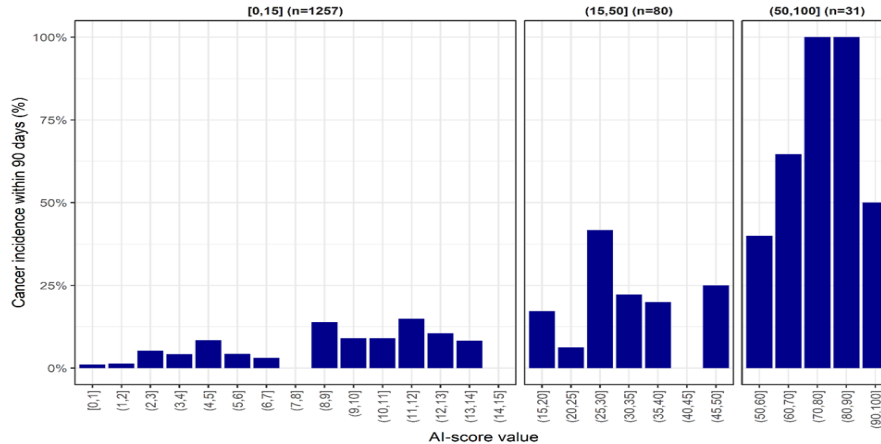
A: Distribution of patients across AI risk scores
B: Observed incidence of cancer across AI risk scores

A



Fordelingen af patienter på tværs af AI -risikoscorer

B



Forekomsten af kræft diagnosticeret inden for 90 dage på tværs af AI -risikoscorer



Præstation af AI-score til at forudsige kræft inden for 90 dage ved forskellige tærskler i Validering Test Cohort II

Cut-off	Cohort number (%)	Cancer number (%)	TP	TN	FP	FN	Sensitivity (%)	Specificity (%)	PPV(%)	NPV(%)
Threshold for negative test										
≤ 1	353(25.8)	4(1.1)	80	349	935	4	95.2	27.2	7.9	98.9
≤ 2	576(42.1)	7(1.2)	77	569	715	7	91.7	44.3	9.7	98.8
≤ 3	726(53.1)	15(2.1)	69	711	573	15	82.1	55.4	10.7	97.9
≤ 4	845(61.8)	20(2.4)	64	825	459	20	76.2	64.3	12.2	97.6
≤ 5	928(67.8)	27(2.9)	57	901	383	27	67.9	70.2	13.0	97.1
≤ 6	997(72.9)	30(3.0)	54	967	317	30	64.3	75.3	14.6	97.0
≤ 7	1061(77.6)	32(3.0)	52	1029	255	32	61.9	80.1	16.9	97.0
≤ 8	1104(80.7)	32(2.9)	52	1072	212	32	61.9	83.5	19.7	97.1
≤ 9	1140(83.3)	37(3.2)	47	1103	181	37	56.0	85.9	20.6	96.8
≤ 10	1173(85.7)	40(3.4)	44	1133	151	40	52.4	88.2	22.6	96.6
≤ 11	1195(87.4)	42(3.5)	42	1153	131	42	50.0	89.8	24.3	96.5
≤ 12	1215(88.8)	45(3.7)	39	1170	114	45	46.4	91.1	25.5	96.3
≤ 13	1234(90.2)	47(3.8)	37	1187	97	47	44.0	92.4	27.6	96.2
≤ 14	1246(91.1)	48(3.9)	36	1198	86	48	42.9	93.3	29.5	96.1
≤ 15	1257(91.9)	48(3.8)	36	1209	75	48	42.9	94.2	32.4	96.2
≤ 16	1262(92.3)	51(4.0)	33	1211	73	51	39.3	94.3	31.1	96.0
≤ 17	1270(92.8)	51(4.0)	33	1219	65	51	39.3	94.9	33.7	96.0
≤ 18	1277(93.3)	51(4.0)	33	1226	58	51	39.3	95.5	36.3	96.0
≤ 19	1280(93.6)	51(4.0)	33	1229	55	51	39.3	95.7	37.5	96.0
≤ 20	1286(94)	53(4.1)	31	1233	51	53	36.9	96.0	37.8	95.9
≤ 25	1302(95.2)	54(4.1)	30	1248	36	54	35.7	97.2	45.5	95.9
≤ 50	1337(97.7)	63(4.7)	21	1274	10	63	25.0	99.2	67.7	95.3
≤ 75	1362(99.6)	79(5.8)	5	1283	1	79	6.0	99.9	83.3	94.2

Cut-off	Cohort number (%)	Cancer number (%)	TP	TN	FP	FN	Sensitivity (%)	Specificity (%)	PPV(%)	NPV(%)
Threshold for positive test										
> 1	1015(74.2)	80(7.9)	80	349	935	4	95.2	27.2	7.9	98.9
> 2	792(57.9)	77(9.7)	77	569	715	7	91.7	44.3	9.7	98.8
> 3	642(46.9)	69(10.7)	69	711	573	15	82.1	55.4	10.7	97.9
> 4	523(38.2)	64(12.2)	64	825	459	20	76.2	64.3	12.2	97.6
> 5	440(32.2)	57(13.0)	57	901	383	27	67.9	70.2	13.0	97.1
> 6	371(27.1)	54(14.6)	54	967	317	30	64.3	75.3	14.6	97.0
> 7	307(22.4)	52(16.9)	52	1029	255	32	61.9	80.1	16.9	97.0
> 8	264(19.3)	52(19.7)	52	1072	212	32	61.9	83.5	19.7	97.1
> 9	228(16.7)	47(20.6)	47	1103	181	37	56.0	85.9	20.6	96.8
> 10	195(14.3)	44(22.6)	44	1133	151	40	52.4	88.2	22.6	96.6
> 11	173(12.6)	42(24.3)	42	1153	131	42	50.0	89.8	24.3	96.5
> 12	153(11.2)	39(25.5)	39	1170	114	45	46.4	91.1	25.5	96.3
> 13	134(9.8)	37(27.6)	37	1187	97	47	44.0	92.4	27.6	96.2
> 14	122(8.9)	36(29.5)	36	1198	86	48	42.9	93.3	29.5	96.1
> 15	111(8.1)	36(32.4)	36	1209	75	48	42.9	94.2	32.4	96.2
> 16	106(7.7)	33(31.1)	33	1211	73	51	39.3	94.3	31.1	96.0
> 17	98(7.2)	33(33.7)	33	1219	65	51	39.3	94.9	33.7	96.0
> 18	91(6.7)	33(36.3)	33	1226	58	51	39.3	95.5	36.3	96.0
> 19	88(6.4)	33(37.5)	33	1229	55	51	39.3	95.7	37.5	96.0
> 20	82(6.0)	31(37.8)	31	1233	51	53	36.9	96.0	37.8	95.9
> 25	66(4.8)	30(45.5)	30	1248	36	54	35.7	97.2	45.5	95.9
> 50	31(2.3)	21(67.7)	21	1274	10	63	25.0	99.2	67.7	95.3
> 75	6(0.4)	5(83.3)	5	1283	1	79	6.0	99.9	83.3	94.2



Eksempel for anvendelse af AI score

Cut-off	Cohort number (%)	Cancer number (%)	TP	TN	FP	FN	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)
Threshold for negative test										
≤ 1	353(25.8)	4(1.1)	80	349	935	4	95.2	27.2	7.9	98.9
≤ 2	576(42.1)	7(1.2)	77	569	715	7	91.7	44.3	9.7	98.8
≤ 3	726(53.1)	15(2.1)	69	711	573	15	82.1	55.4	10.7	97.9
≤ 4	845(61.8)	20(2.4)	64	825	459	20	76.2	64.3	12.2	97.6
≤ 5	928(67.8)	27(2.9)	57	901	383	27	67.9	70.2	13.0	97.1

Cut-off	Cohort number (%)	Cancer number (%)	TP	TN	FP	FN	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)
>5	440(32.2)	57(13.0)	57	901	383	27	67.9	70.2	13.0	97.1
>6	371(27.1)	54(14.6)	54	967	317	30	64.3	75.3	14.6	97.0
>7	307(22.4)	52(16.9)	52	1029	255	32	61.9	80.1	16.9	97.0
>8	264(19.3)	52(19.7)	52	1072	212	32	61.9	83.5	19.7	97.1
>9	228(16.7)	47(20.6)	47	1103	181	37	56.0	85.9	20.6	96.8
>10	195(14.3)	44(22.6)	44	1133	151	40	52.4	88.2	22.6	96.6
>11	173(12.6)	42(24.3)	42	1153	131	42	50.0	89.8	24.3	96.5
>12	153(11.2)	39(25.5)	39	1170	114	45	46.4	91.1	25.5	96.3
>13	134(9.8)	37(27.6)	37	1187	97	47	44.0	92.4	27.6	96.2
>14	122(8.9)	36(29.5)	36	1198	86	48	42.9	93.3	29.5	96.1
>15	111(8.1)	36(32.4)	36	1209	75	48	42.9	94.2	32.4	96.2
>16	106(7.7)	33(31.1)	33	1211	73	51	39.3	94.3	31.1	96.0
>17	98(7.2)	33(33.7)	33	1219	65	51	39.3	94.9	33.7	96.0
>18	91(6.7)	33(36.3)	33	1226	58	51	39.3	95.5	36.3	96.0
>19	88(6.4)	33(37.5)	33	1229	55	51	39.3	95.7	37.5	96.0
>20	82(6.0)	31(37.8)	31	1233	51	53	36.9	96.0	37.8	95.9
>25	66(4.8)	30(45.5)	30	1248	36	54	35.7	97.2	45.5	95.9
>50	31(2.3)	21(67.7)	21	1274	10	63	25.0	99.2	67.7	95.3
>75	6(0.4)	5(83.3)	5	1283	1	79	6.0	99.9	83.3	94.2

Konklusion

- *AI-model baseret på rutinemæssige blodprøver - i stand til at give en brugervenlig risikoscore til at forudsige kræft inden for 90 dage.*
- *AI-risikoscoren kan vise sig at være et værdifuldt værktøj i den kliniske beslutningstagning*
 - *understøtter den praktiserende læge, om en patient har brug for hurtigere yderligere undersøgelser eller i stedet en vågent ventestrategi.*
- *AI -score skal dog yderligere valideres eksternt - anvendelighed i andre populationer.*



Perspektiver

- *Fremtidig forbedring af AI -risikoscoren ved yderligere udvikling og tilpasning af analyserne inkluderet i blodprøveprofilen*
- *Derudover ved at kombinere testresultater og demografiske data i fremtidige forudsigelsesmodeller.*



TAK FOR OPMÆRKSOMHED!



Patricia.Diana.Sorensen@rsyd.dk